DOI: 10.53469/jpce.2025.07(08).06

A VMD-Based Method for Outlier Correction in Structural Health Monitoring Data

Chen Zhao*, Xin Gao

College of Construction Engineering, Jilin University, Changchun 130026, China
**Correspondence Author, zhaoc23@mails.jlu.edu.cn*

Abstract: ***The quality of Structural Health Monitoring (SHM) data is paramount to the accuracy of structural condition assessment and service life prediction. However, monitoring data acquired in the field often exhibit significant non-stationarity and contain outliers due to environmental interference and other factors, posing severe challenges to subsequent data analysis. Traditional outlier detection methods often suffer from low accuracy and high false-positive rates when processing non-stationary signals, owing to interference from trend and periodic components. To address this issue, this study proposes a joint data correction framework based on Variational Mode Decomposition (VMD) and the Isolation Forest algorithm. The proposed method first utilizes the adaptive decomposition capability of VMD to decompose the original non-stationary signal into a series of Intrinsic Mode Functions (IMFs). The first component (IMF1) is extracted as the macroscopic trend of the signal, achieving efficient signal detrending. Subsequently, the Isolation Forest algorithm is applied to the detrended residual signal to accurately identify and locate outliers. Finally, linear interpolation is employed to correct the identified outliers. To validate the effectiveness of the proposed method, a synthetic dataset comprising trend, multi-periodic oscillations, and noise was constructed. Comparative experimental results demonstrate that the proposed VMD-Isolation Forest framework significantly enhances the accuracy and robustness of outlier detection compared to the direct application of the 3σ rule or the Isolation Forest algorithm. It effectively corrects anomalous disturbances while maximally preserving the intrinsic dynamic characteristics of the original signal. This research provides an efficient and reliable preprocessing paradigm for non-stationary SHM data, laying a solid data foundation for subsequent high-precision structural damage identification and performance prediction models.***

Keywords: Structural Health Monitoring (SHM), Outlier Detection, Variational Mode Decomposition (VMD), Isolation Forest, Non-stationary Signal, Data Preprocessing**.**

1. Introduction

Structural Health Monitoring (SHM) is a core technology for ensuring the operational safety of critical infrastructure. By collecting real-time dynamic response data of structures (e.g., stress, vibration, and displacement), SHM provides a crucial basis for assessing the evolution of structural damage and predicting remaining service life (Farrar and Worden 2012). With the rapid development of complex engineering projects such as large-scale bridges and long-span spatial structures, traditional monitoring methods relying on manual inspection and preset threshold alarms are no longer adequate for modern engineering applications due to their inherent limitations in monitoring accuracy and timeliness (Giordano, Quqa et al. 2023, Qi, Hou et al. 2024). In this context, efficient decomposition theories and precise modeling methods for non-stationary signals have become central research topics in the SHM field (Dong, Li et al. 2010, Vazirizade, Bakhshi et al. 2018, Bisheh and Amiri 2023).

Early signal analysis methods primarily relied on the Fourier Transform. However, this method is only applicable to stationary signals and struggles to effectively capture and analyze non-stationary signals with time-varying characteristics. Although Wavelet Analysis partially addresses this issue through its multi-scale time-frequency decomposition capabilities, the selection of its basis functions is highly dependent on a priori knowledge, and it suffers from inherent drawbacks such as energy leakage (Mallat 2008). Subsequently, the Empirical Mode Decomposition (EMD) method, proposed by Huang et al. (1998), became a mainstream technique in the field due to its excellent adaptive properties. However, the mode mixing phenomenon inherent in EMD—where frequency components of different physical significance are coupled within the same mode—severely compromises the reliability of its decomposition results (Huang, Shen et al. 1998). For instance, when processing signals containing high-frequency transient noise, EMD is prone to generating spurious modal components, which can lead to a misjudgment rate of over 30% in subsequent structural damage identification (Lei, Lin et al. 2013).

To overcome these limitations, Dragomiretskiy et al. (2014) introduced a novel signal processing model called Variational Mode Decomposition (VMD), which features adaptivity, quasi-orthogonality, and a completely non-recursive nature. Its core idea is to transform the complex signal decomposition process into a mathematical problem of finding the optimal solution to an unconstrained variational problem. Consequently, it is frequently applied to the analysis and processing of non-stationary signals (Li, Liu et al. 2022). Li et al. (Li, Zou et al. 2023) proposed a progressive decomposition and double-screening strategy based on VMD to enhance the extraction of weak fault features in machinery, demonstrating its ability to accurately identify early mechanical faults even under noise interference, thus providing new insights for complex SHM. Ma et al. (Ma, Wang et al. 2022) utilized VMD to extract multi-scale features from non-stationary load signals in power systems and combined it with a Recurrent Neural Network (RNN) for classification, achieving an identification accuracy of 89.8% on a dataset of 200,000 users. Du (Du 2022) combined VMD with LSTM for damage prediction in concrete frames and long-span structures, optimizing the sensitivity and generalization of damage identification by decomposing structural response signals with VMD and inputting the multi-frequency features into an LSTM network.

Due to the susceptibility of structures to seasonal climate and environmental damage, SHM data often contain a significant number of outliers, which adversely affect subsequent data modeling and prediction. Traditional statistical methods for outlier detection, such as the 3σ rule and box plot analysis, are widely used due to their simplicity, but they are limited in their effectiveness at identifying complex patterns of anomalies or non-Gaussian noise. As research has progressed, machine learning-based anomaly detection methods, including clustering-based algorithms (e.g., K-means), classification-based algorithms (e.g., Support Vector Machine, SVM), and reconstruction error-based algorithms (e.g., Principal Component Analysis, PCA; Autoencoder), have shown better adaptability in SHM data anomaly identification. Recently, deep learning anomaly detection algorithms specifically designed for time-series data, such as LSTM-based models, have also been applied in the SHM domain. Furthermore, signal decomposition techniques like VMD can assist in identifying anomalous components by analyzing the characteristics (e.g., energy, frequency mutations) of the decomposed Intrinsic Mode Functions (IMFs).

To address the aforementioned issues, this study designs a VMD-based outlier detection method: (1) Decompose the sensor-collected data using VMD and extract the results; (2) Use the resulting IMF1 component as the data's trend component and subtract it from the original data to obtain the detrended data; (3) Perform outlier detection on the detrended data, mark the detected outliers as missing values, and then impute them.

2. Data Preprocessing Workflow

Given the potential for outliers in monitoring data, and since sensor-collected data are typically non-stationary time series containing significant trend and complex periodic components, the performance of traditional anomaly detection methods is easily compromised when applied directly to such data. This interference can lead to a significant increase in detection errors. Therefore, this study proposes a strategy that combines VMD for data detrending with various outlier detection methods.

The strategy first decomposes the original signal $x(t)$ using VMD to obtain a set of Intrinsic Mode Functions (IMFs) $\{IMF_i(t)\}_{i=1}^{n}$. The first IMF component, IMF1, is selected as the low-frequency trend of the original signal, primarily because IMF1 can effectively represent the inherent frequency information and long-term evolutionary trend of the structural response. Subsequently, this low-frequency trend is separated from the original signal using the following equation to obtain the detrended signal $x_{detrended}(t)$:

$$x_{trend}(t) = IMF_1(t), \quad x_{detrended}(t) = x(t) - x_{trend}(t) \tag{1}$$

The Isolation Forest algorithm (Liu, Ting et al. 2008) and the 3σ rule are then applied separately to the detrended signal $x_{detrended}(t)$ to detect outliers. Once an outlier is detected, it is temporarily marked as a missing value, which is subsequently imputed. This process is repeated until no new outliers are detected or a preset iteration limit is reached.

For the data points marked as missing values, this study employs the linear interpolation method for imputation (Blu, Thevenaz et al. 2004). The mathematical expression is as follows:

$$y(t) = y(t_{prev}) + \frac{y(t_{next}) - y(t_{prev})}{t_{next} - t_{prev}} \cdot (t - t_{prev}) \tag{2}$$

where t_{prev} and t_{next} are the timestamps of the nearest valid data points before and after the missing point, respectively. This method effectively preserves the local trend characteristics of the original data and, compared to traditional mean-filling methods, can more effectively suppress the bias that may be introduced by smoothing.

3. Experiment

3.1 Data Characteristics

To validate the effectiveness of the proposed method, this study constructed a synthetic dataset with non-stationary characteristics, including a trend component, multi-frequency periodic oscillation components, and Gaussian noise (Bandara, Hyndman et al. 2021). The synthetic dataset simulates the monitoring signals from 10 virtual stress sensors over 30 consecutive days, with a sampling interval of 30 minutes. The mathematical expression for the synthetic time-series data is defined as follows:

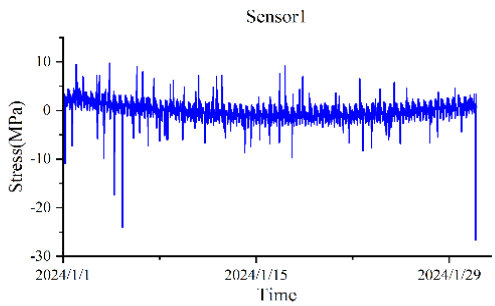
$$X_t(i) = T_t + S_t(i) + \gamma R_t \tag{3}$$

where the periodic term $S_t(i)$ is a linear superposition of three Fourier series with different periods, and each Fourier series contains 5 pairs of random coefficients that are independently and identically distributed according to a standard normal distribution. The residual term R_t consists of Gaussian noise with a mean of 0 and a standard deviation of 1, simulating random interference during the measurement process. A weight of γ =0.2 to this term. The trend term Tt is defined by a quadratic polynomial:

$$T_t = N_1(t + \frac{n}{2}(N_2 - 1))^2 \tag{4}$$

where N_1 and N_2 are independent random variables following a N(0,1) distribution.

To further test the robustness and effectiveness of the proposed data preprocessing method, a certain proportion and magnitude of outliers were intentionally injected into the generated synthetic data. Figure 1 visually displays the original time-series data and the distribution of outliers for three representative sensors (numbered 1-3) selected from the 10 virtual sensors. It can be clearly observed from the figure that these original signals exhibit significant non-stationarity, and the complex trend and multi-frequency periodic oscillation components are highly coupled.



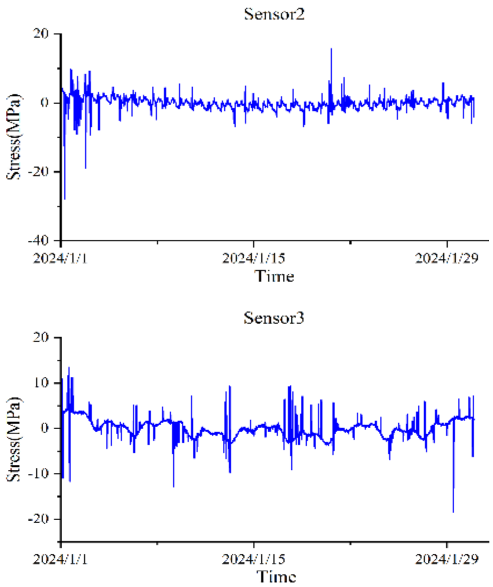
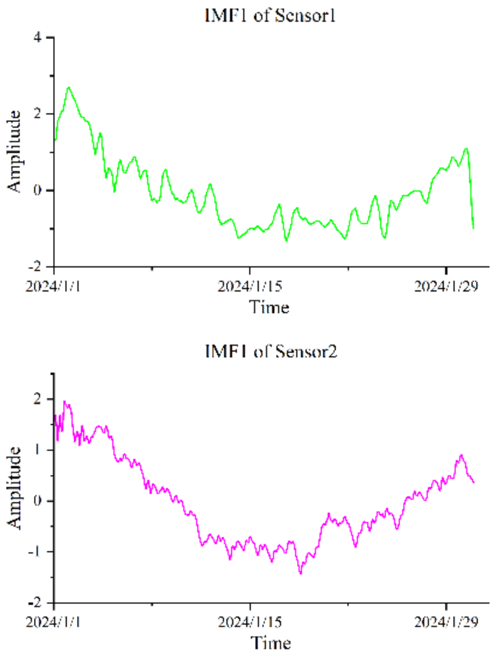


Figure 1: Original time-series plots for three sensors (Sensor 1, Sensor 2, Sensor 3)

3.2 Outlier Correction

Following the preprocessing workflow described in Section 2, the original signal x(t) is first decomposed using VMD. After decomposition, the low-frequency first intrinsic mode function, IMF1, is extracted as the low-frequency trend component $x_{trend}(t)$ (Figure 2). Subsequently, this trend component is subtracted from the original signal to reconstruct the detrended signal $x_{detrended}(t)$ (Figure 3). The Isolation Forest algorithm and the 3σ algorithm are then applied separately to the detrended signal $x_{detrended}(t)$ for outlier detection. By comparing the original signals in Figure 1 with the detrended signals in Figure 3, it can be observed that the VMD detrending process effectively eliminates the prominent trend interference in the original data. This preprocessing step allows subsequent anomaly detection to focus more accurately on the local random fluctuation characteristics of the signal itself, rather than being misled by trend variations.



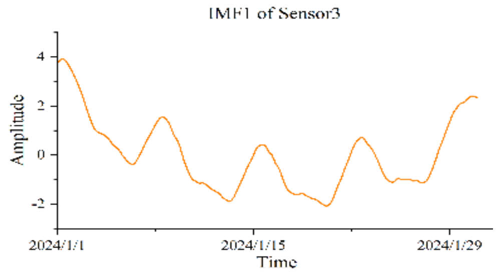


Figure 2: Low-frequency trend components $x_{trend}(t)$ (IMF1) for three sensors

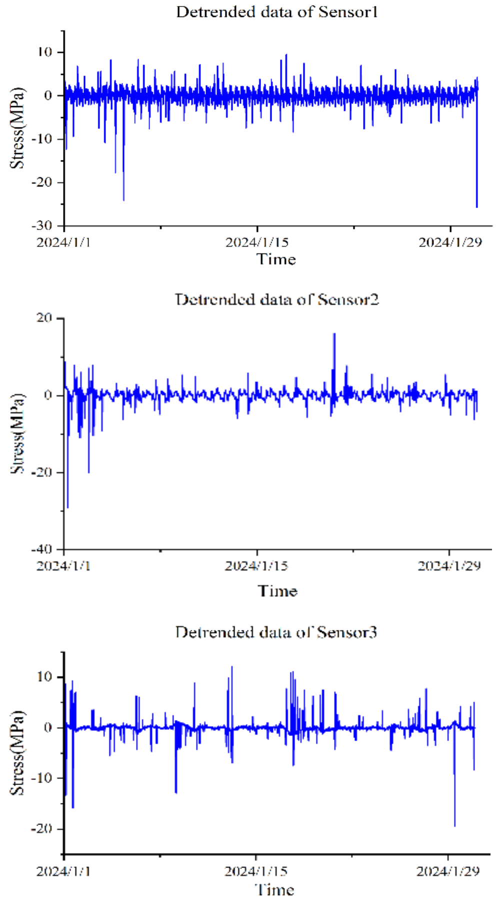
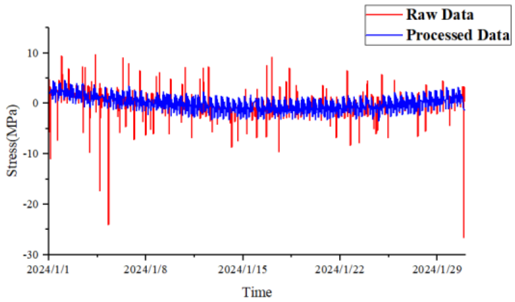


Figure 3: Detrended signals $x_{detrended}(t)$ for three sensors

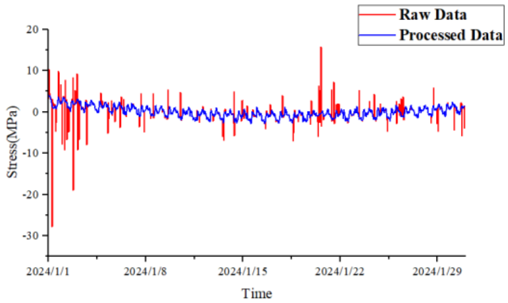
3.3 Preprocessing Results and Comparison

To quantitatively evaluate the effectiveness of the proposed joint preprocessing framework, this section provides a comparative analysis of different preprocessing strategies. The primary objectives are to verify the necessity of VMD-based detrending for improving outlier detection accuracy and to compare the performance of different detection algorithms (Isolation Forest vs. 3σ rule) on the detrended signals.

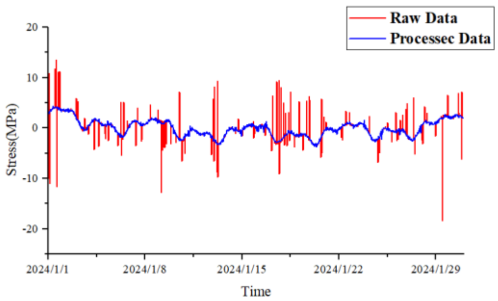
Figures 4(a)-(c) show a comparison between the data corrected by our proposed VMD-Isolation Forest joint preprocessing method and the original data. It is clearly observable that the abrupt changes, spikes, and other anomalies present in the original signal have been effectively identified and smoothed. Crucially, this correction process accurately preserves the overall long-term trend and intrinsic periodic oscillation features of the signal while eliminating local anomalous disturbances, without introducing significant signal distortion. This preliminarily validates the superior balance of the joint method between effective data anomaly repair and signal fidelity preservation.



(a) Comparison for Sensor1



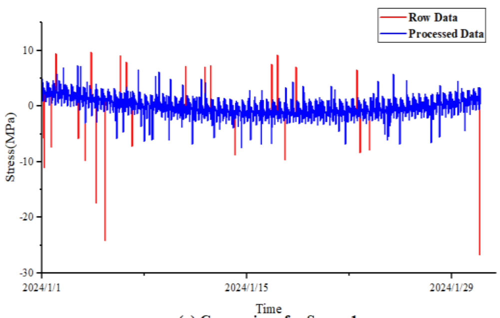
(b) Comparison for Sensor2



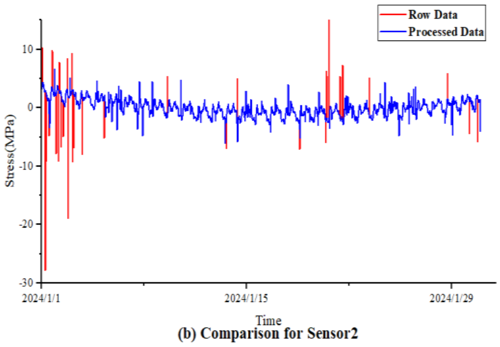
(c) Comparison for Sensor3

Figure 4: Comparison of original data with data corrected by the VMD-Isolation Forest method

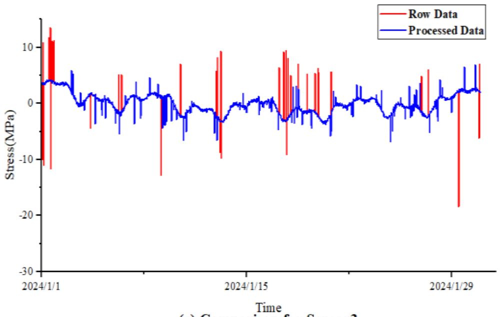
To further highlight the advantages of the proposed method, Figures 5(a)-(c) display the results of the VMD-3σ method. Although this method also corrects some outliers, its effectiveness is slightly inferior to the VMD-Isolation Forest method, especially when dealing with boundary or asymmetrically distributed anomalies, where it may exhibit missed detections. This indicates that for residual sequences presenting complex non-Gaussian distributions after detrending, the unsupervised learning-based Isolation Forest algorithm possesses greater robustness and adaptability than the 3σ rule, which relies on the assumption of a normal distribution.



(a) Comparison for Sensor1



(b) Comparison for Sensor2

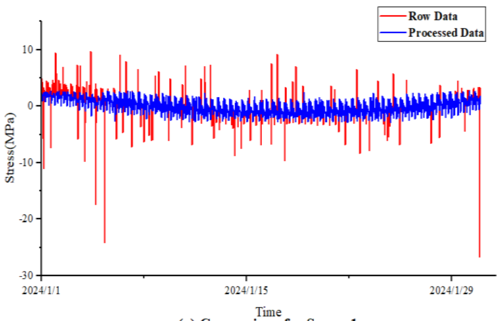


(c) Comparison for Sensor3

Figure 5: Comparison of original data with data corrected by the VMD-3σ method

Furthermore, to demonstrate the necessity of VMD detrending as the primary preprocessing step, we conducted a control experiment by directly applying outlier detection algorithms to the original signals. Figures 6(a)-(c) show the results of directly using the Isolation Forest algorithm. Due to the strong non-stationarity of the original signal, its inherent trend and periodic components severely interfered with the algorithm's density estimation. The results show that this method not only failed to identify all anomalies (as shown in Figure 6(b)) but also incorrectly classified some normal data points as outliers during the initial phase of the signal due to rapid trend changes (as shown in Figures 6(a) and 6(c)), leading to "false positives."

Figures 7(a)-(c) display the results of directly applying the 3σ rule, which performed even more poorly. The trend and periodic components of the original signal "contaminated" the overall mean and variance, causing the calculated outlier detection threshold to be excessively large. Consequently, this method failed to identify almost any of the true outliers, proving that without detrending, traditional statistical methods are essentially ineffective for anomaly detection in non-stationary time-series data.



(a) Comparison for Sensor1

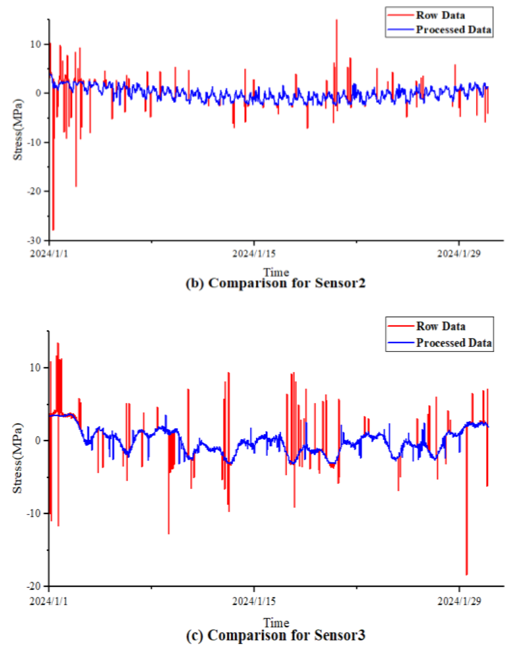


Figure 6: Comparison of original data with data corrected directly by the Isolation Forest method

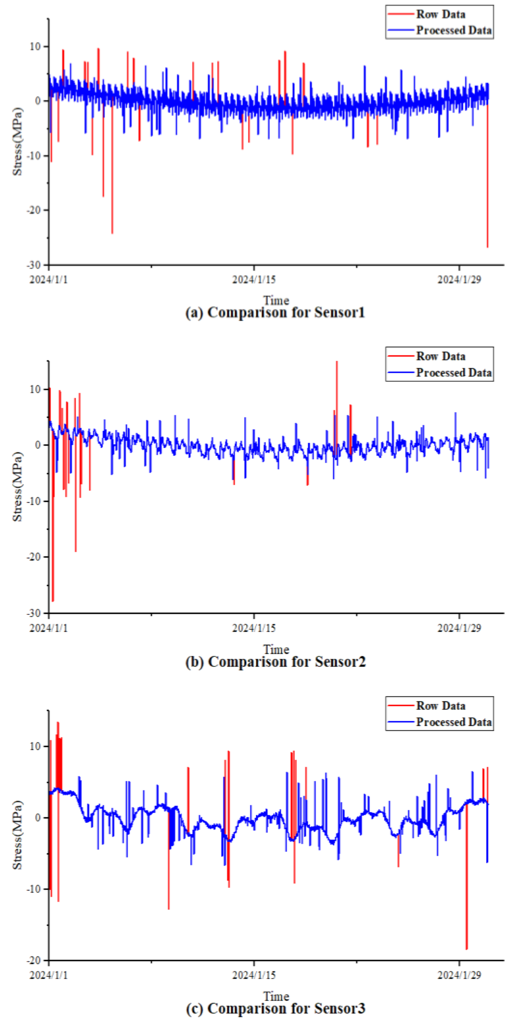


Figure 7: Comparison of original data with data corrected directly by the 3σ rule

In summary, the analysis of the four sets of comparative experiments leads to the following conclusions:

1) **VMD Detrending is a Critical Step:** Performing outlier detection directly on non-stationary raw data, regardless of the algorithm used, is highly susceptible to interference from trend and periodic components, leading to a significant decline in detection performance, or even failure.

2) **Isolation Forest Algorithm Exhibits Superior Performance:** After VMD detrending, the Isolation Forest algorithm demonstrates more accurate and robust identification of outliers in complex data compared to the traditional 3σ rule.

3) Therefore, the VMD-Isolation Forest joint preprocessing framework proposed in this study can most effectively correct anomalies in SHM monitoring data, providing a high-quality data foundation for subsequent high-precision time-series prediction models.3.

4. Conclusion and Outlook

4.1 Conclusion

To address the prevalent issues of non-stationarity and outlier contamination in structural health monitoring data, this study has proposed and validated a novel data preprocessing framework that combines Variational Mode Decomposition (VMD) and the Isolation Forest algorithm. Through systematic experimental analysis on synthetic time-series data containing complex trends, periodic components, and injected anomalies, the following main conclusions are drawn:

1) **Necessity and Effectiveness of VMD Detrending:** The experimental results unequivocally confirm that applying outlier detection algorithms (whether statistical or machine learning-based) directly to non-stationary raw signals leads to performance degradation due to severe interference from the signal's intrinsic trend and periodic components, resulting in numerous missed detections and false positives. Extracting and separating the signal's low-frequency trend component via VMD effectively eliminates this interference and is a critical prerequisite for achieving accurate outlier detection.

2) **Superiority of the Isolation Forest Algorithm:** When detecting anomalies in the detrended residual signal, the unsupervised learning-based Isolation Forest algorithm demonstrates greater robustness and higher accuracy compared to the traditional 3σ rule. Particularly in handling asymmetric or complexly distributed anomalies, Isolation Forest can more effectively identify true outliers, thereby enhancing the reliability of the correction.

3) **Comprehensive Performance of the Joint Framework:** The proposed VMD-Isolation Forest joint preprocessing framework successfully integrates the advantages of both methods. It not only accurately identifies and smooths abrupt changes and spikes in the data but also preserves the long-term trends and periodic oscillation features, which are crucial for structural analysis, thus avoiding signal distortion during the correction process

In summary, the method proposed in this study provides an efficient and reliable data-cleaning tool for the SHM field, capable of significantly improving the quality of monitoring data and providing high-quality input for subsequent advanced analysis tasks such as time-series forecasting, damage identification, and remaining service life assessment.

4.2 Outlook

Although this study has validated the effectiveness of the proposed method using synthetic data, there is still room for further in-depth research and expansion. Future work can be pursued in the following areas:

1) **Validation and Application on Real Engineering Data:** Apply the method to long-term monitoring data from real engineering structures (e.g., large bridges, high-rise buildings, or offshore platforms). In real-world scenarios, data characteristics are more complex, potentially containing various types of noise and anomalies from different sources. This will serve as the ultimate test of the method's generalization ability and robustness.

2) **Adaptive Optimization of VMD Parameters:** The decomposition performance of VMD depends on the selection of key parameters such as the number of modes (K) and the penalty factor (α). Currently, these parameters are often set based on experience. Future research could explore the integration of intelligent optimization algorithms (e.g., Genetic Algorithms, Particle Swarm Optimization, or Bayesian Optimization) to achieve adaptive parameter tuning, thereby further enhancing the accuracy and automation of signal decomposition.

3) **Exploration of Fusion with Other Anomaly Detection Algorithms:** While Isolation Forest has shown excellent performance, it is worthwhile to explore combining VMD detrending with other advanced deep learning anomaly detection models (e.g., models based on Autoencoders or Long Short-Term Memory networks - LSTM) to assess whether further performance gains can be achieved in specific application scenarios.

4) **Quantitative Evaluation of Impact on Downstream Analysis Tasks:** Systematically investigate the improvement in the accuracy of specific downstream tasks (e.g., structural modal parameter identification, damage localization and quantification, crack propagation prediction) using data preprocessed by this method. Through quantitative analysis, the practical value of this preprocessing method within the entire SHM workflow can be more comprehensively revealed.

References

[1] Bandara, K., R. J. Hyndman and C. Bergmeir (2021). "MSTL: A Seasonal-Trend Decomposition Algorithm for Time Series with Multiple Seasonal Patterns." https://doi.org/10.48550/arXiv.2107.13462

[2] Focus to learn more.

[3] Bisheh, H. B. and G. G. Amiri (2023). "Structural damage detection based on variational mode decomposition and kernel PCA-based support vector machine." Engineering Structures **278**: 115565.

[4] Blu, T., P. Thevenaz and M. Unser (2004). "Linear interpolation revitalized." IEEE Transactions on Image Processing **13**(5): 710-719.

[5] Dong, Y., Y. Li and M. Lai (2010). "Structural damage detection using empirical-mode decomposition and vector autoregressive moving average model." Soil Dynamics and Earthquake Engineering **30**(3): 133-145.

[6] Du, R. (2022). Optimization of CNN and LSTM Based Application on RC Frame and Long-Span Structural Health Monitoring. CONF-SPML.

[7] Farrar, C. R. and K. Worden (2012). Structural Health Monitoring: A Machine Learning Perspective, John Wiley & Sons, Ltd.

[8] Giordano, P. F., S. Quqa and M. P. Limongelli (2023). "The value of monitoring a structural health monitoring system." Structural Safety **100**: 102280.

[9] Huang, N. E., Z. Shen, S. R. Long, M. L. C. Wu, H. H. Shih, Q. N. Zheng, N. C. Yen, C. C. Tung and H. H. Liu (1998). "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis." Proceedings of the Royal Society a-Mathematical Physical and Engineering Sciences **454**(1971): 903-995.

[10] Lei, Y., J. Lin, Z. He and M. J. Zuo (2013). "A review on empirical mode decomposition in fault diagnosis of rotating machinery." Mechanical Systems and Signal Processing **35**(1): 108-126.

[11] Li, C., Y. Liu, Y. Liao and J. Wang (2022). "A VME method based on the convergent tendency of VMD and its application in multi-fault diagnosis of rolling bearings." Measurement **198**.

[12] Li, Y., L. Zou, C.-G. Lee and F. Xu (2023). "A progressive decomposing and double screening strategy of VMD for weak fault extraction of hoisting machinery." Structural Health Monitoring-an International Journal **22**(2): 1140-1160.

[13] Lim, S.-C., J.-H. Huh, S.-H. Hong, C.-Y. Park and J.-C. Kim (2022). "Solar Power Forecasting Using CNN-LSTM Hybrid Model." Energies **15**(21).

[14] Liu, F. T., K. M. Ting and Z. H. Zhou (2008). Isolation Forest. 2008 Eighth IEEE International Conference on Data Mining.

[15] Ma, Y., H. Wang, J. Yang, Z. Chen, J. Yuan and M. O. Okoye (2022). "RNN-Assisted Feature-Extraction VMD for Load Classification in Cloud Computing Platform." Journal of Circuits Systems and Computers **31**(06).

[16] Mallat, S. (2008). A Wavelet Tour of Signal Processing, Academic Press.

[17] Mirjalili, S. and A. Lewis (2016). "The Whale Optimization Algorithm." Advances in Engineering Software **95**: 51-67.

[18] Qi, T., Z. Hou and L. Yu (2024). "Structural damage identification based on dual sensitivity analysis from optimal sensor placement." Journal of Infrastructure Intelligence and Resilience **3**(3): 100110.

[19] Sharma, S. and S. Sen (2023). "Real-time structural damage assessment using LSTM networks: regression and classification approaches." Neural Computing and Applications **35**(1): 557-572.

[20] Tong, X., J. Wang, C. Zhang, T. Wu, H. Wang and Y. Wang (2022). "LS-LSTM-AE: Power load forecasting via Long-Short series features and LSTM-Autoencoder." Energy Reports **8**: 596-603.

[21] Vazirizade, S. M., A. Bakhshi and O. Bahar (2018). Structural Damage Detection Using Ensemble Empirical

Mode Decomposition, Hilbert Transform and Artificial Neural Networks.

[22] Wang, J. and S. Guo (2020). "Application of adaptive VMD algorithm in rolling bearing failure diagnosis." Electromechanical Engineering Technology **49**(11): 161-164.

[23] Wen, Y., S. Pan, X. Li and Z. Li (2024). "Highly fluctuating short-term load forecasting based on improved secondary decomposition and optimized VMD." Sustainable Energy, Grids and Networks **37**: 101270.

[24] Yan, X., M. Jia and L. Xiang (2016). "Compound fault diagnosis of rotating machinery based on OVMD and a 1.5-dimension envelope spectrum." Measurement Science and Technology **27**(7).

[25] Yang, J., K. Yan, Z. Wang and X. Zheng (2022). "A Novel Denoising Method for Partial Discharge Signal Based on Improved Variational Mode Decomposition." Energies **15**(21).

[26] Yang, K., Y. Ding, F. Geng, H. Jiang and Z. Zou (2023). "A multi-sensor mapping Bi-LSTM model of bridge monitoring data based on spatial-temporal attention mechanism." Measurement **217**: 113053.

[27] Yue, Z., Y. Ding, H. Zhao and Z. Wang (2022). "Mechanics-Guided optimization of an LSTM network for Real-Time modeling of Temperature-Induced deflection of a Cable-Stayed bridge." Engineering Structures **252**: 113619.

[28] Zhang, Z., H. Li, G. Meng, X. Tu and C. Cheng (2016). "Chatter detection in milling process based on the energy entropy of VMD and WPD." International Journal of Machine Tools & Manufacture **108**: 106-112.

[29] Zhao, G., K. Xing, Y. Wang, H. Qian and M. Zhang (2023). "Long Short-Term Memory Network for Predicting Wind-Induced Vibration Response of Lightning Rod Structures." Buildings **13**(5): 1256.